When the Comparison Is the Problem: Spatial Resolution and Validation Bias in InSAR-Derived Coastal Subsidence Assessments Along the U.S. Gulf Coast

Manoochehr Shirzaei[1,2,3*], Leonard Ohenhen[4], Carmen Atkins[1], Oluwaseyi Dasho[1], Nitheshnirmal Sadhasivam[1], Nivedita P. Kamaraj[1], Florence Onyike[1], Olasunkanmi Olorunsaye[1], Esther O. Oyedele[1] and Susanna Werth[1,2]

[1]Department of Geosciences, Virginia Tech, Blacksburg, VA, USA
[2]United Nations University Institute for Water, Environment and Health, Richmond Hill, ON, Canada
[3]UNESCO Land Subsidence International Initiative (laSII), Place de Fontenoy 7, Paris, France
[4]Department of Earth System Science, University of California, Irvine, Irvine, CA, USA
[*]Correspondence: shirzaei@vt.edu

Abstract

Li et al. (2026) compare two InSAR-derived surface-elevation change datasets for the central U.S. Gulf Coast and conclude that InSAR is unreliable in vegetated coastal settings for rates below 5 mm/yr. While InSAR reproducibility is a timely and consequential question, we demonstrate that the paper's principal conclusions rest on three methodological decisions that critically undermine the comparison: (1) spatial aggregation of the O24 dataset from its native 50 m to 1 km prior to comparison, a ~400x reduction in pixel density that destroys the sub-kilometer spatial structure for which the dataset was designed; (2) a progressively filtered GNSS validation network of only ~20 stations concentrated in atypical stable Pleistocene upland settings, contrasting with O24's original validation across 157 stations spanning the full coastal domain; and (3) a 5 mm/yr caution threshold derived from inter-product disagreement between two methodologically dissimilar datasets rather than from principled uncertainty quantification. We validate O24 at its native 50 m resolution against 88 GNSS stations from the Nevada Geodetic Laboratory within the Li et al. study domain, obtaining a residual standard deviation of 1.6 mm/yr, consistent with Ohenhen et al. (2024) and directly contradicting the paper's characterization of O24 performance. We call on the InSAR community to prioritize coordinated benchmarking and invest in methodological literacy around resolution, coherence, and uncertainty quantification, so that inter-product disagreement is neither conflated with measurement failure nor permitted to drive policy-relevant conclusions without rigorous independent validation.

1. Introduction

Li et al. (2026) address a genuine and consequential challenge: two recent InSAR-derived SEC datasets for the central U.S. Gulf Coast, Ohenhen et al. (2024; hereafter O24) and Wang et al. (2024; hereafter W24), yield similar regional mean rates yet show negligible pixel-by-pixel spatial correlation ($R^2 = 0.05$) outside urbanized areas. This finding, if interpreted at face value, would have serious implications for the use of InSAR in coastal subsidence monitoring and sea-level-rise impact assessments.

We agree that reproducibility is a critical issue in this field, and we support the paper's call for harmonized processing frameworks and systematic cross-validation. However, the analytical choices underlying the comparison introduce confounds of sufficient magnitude to cast doubt on the paper's primary conclusion, namely, that InSAR is presently unreliable for capturing rates below 5 mm yr^{-1} in vegetated coastal settings. In what follows, we document three specific methodological concerns and identify statements in the paper that are inconsistent with the evidence the authors themselves present.

2. Spatial Resampling of O24: A Non-Neutral Transformation

The comparison between O24 and W24 is not conducted at the native resolution of either dataset. Rather, Li et al. (2026) aggregate O24 from its native 50 m pixel spacing to the 1 km pixel spacing of W24 by computing median SEC values within a 500 m search radius around each W24 centroid. The effect is a ~400-fold reduction in spatial information density: each 1 km² pixel in the resampled product represents the central tendency of approximately 400 original observations, each of which may reflect distinct land-cover types, subsidence drivers, or coherence conditions.

O24 was designed specifically to resolve fine-scale spatial heterogeneity in coastal subsidence, including localized signals from differential sediment compaction, fault-zone deformation, and fluid extraction. Its GNSS-constrained stochastic inversion framework exploits the spatial density of InSAR measurements precisely to differentiate these signals. Collapsing this information to a 1 km grid does not simply reduce resolution; it destroys the sub-kilometer spatial structure that distinguishes O24's design from W24's. The aggregation also mixes land-cover classes that O24 resolves separately, including urban-wetland boundaries that are among the most dynamically and geodetically interesting features in the coastal zone.

The authors justify this procedure by noting that search radii from 100–400 m and 1000 m produced only minor differences. This confirms the aggregation is internally stable, but it does not validate the approach: consistency across aggregation radii simply means the spatial averaging has already erased the sub-kilometer variability at the smallest radius tested. The appropriate comparison would preserve O24 at 50 m resolution while spatially interpolating W24 to the finer grid, or, equivalently, report the correlation as a function of spatial scale to determine at which resolution, if any, the two products agree.

The consequence for the headline result is significant. The $R^2 = 0.05$ figure, the evidentiary cornerstone of the paper's most far-reaching claims, may be partly or substantially an artifact of the asymmetric resampling, not a property of InSAR coherence in vegetated landscapes. Until the comparison is conducted at native resolution, or at a series of spatial scales, this cannot be determined.

3. GNSS Validation Network: Filtering, Scope, and Representativeness

Li et al. (2026) establish a 'background VLM' benchmark by progressively filtering 110 candidate GNSS sites to a final set of 36, of which only approximately 20 fall within the region covered by both InSAR datasets at 1 km resolution. The filtering logic, excluding Holocene-deposit sites, all Texas sites, and stations with elevated fluid extraction activity, is defensible in principle: the goal is to isolate GIA from anthropogenic and compaction signals. In practice, however, the resulting network raises representativeness concerns that the paper does not adequately address.

The exclusion of all 30 Texas GNSS sites is particularly blunt. The stated rationale is that 80% of Texas sites have nearby active wells. The remaining 20%, six sites without proximate extraction activity, located on the Pleistocene surface, are discarded without explanation. A more rigorous approach would apply extraction-volume-weighted or distance-weighted criteria to distinguish sites by degree of anthropogenic influence, rather than imposing a state-level binary exclusion. Retaining even a subset of unaffected Texas stations would extend the network westward and improve spatial coverage across the study domain.

More fundamentally, the 20 sites used to evaluate InSAR performance are concentrated in Pleistocene upland settings, by design, the most stable, least dynamic parts of the landscape. InSAR-derived SEC is most consequential in the adjacent Holocene coastal lowlands, where subsidence is fastest, vegetation is densest, and decorrelation noise is most severe. Its limitations are most relevant to policy. Demonstrating that InSAR does not closely match GNSS at stable upland sites does not constitute a validation failure in the coastal lowland context, where the datasets are primarily intended to be used. The physical processes, signal magnitudes, and noise environments differ substantially between these two settings.

The contrast with O24's own validation framework is instructive. Ohenhen et al. (2024) validated their dataset using 157 GNSS sites distributed across the full U.S. Gulf Coast, including settings spanning the full range of land-cover and subsidence conditions, yielding a standard deviation of 1.5 mm/yr for the difference between O24 and GNSS datasets. The use of a narrow, filtered subset in Li et al. (2026) is not a methodological improvement; it is a restriction of scope that the paper does not adequately acknowledge when generalizing its conclusions.

We recommend that the authors supplement their Pleistocene upland comparison with available Holocene-setting benchmarks, including rod surface-elevation table–marker horizon (RSET-MH) records, continuously operating tide gauge records, and additional GNSS sites, to assess whether the conclusions hold across the full range of coastal environments relevant to the paper's claims.

4. The 5 mm yr^{-1} Caution Threshold: Derivation and Internal Consistency

Li et al. (2026) recommend that InSAR-derived vertical velocities below 5 mm yr^{-1} be "interpreted with utmost caution" in regions lacking dense GNSS ground-truth. This threshold is derived from the 95th percentile of absolute SEC differences between O24 and W24 in medium-to-high-developed urban areas, the land-cover class where InSAR coherence is most reliable and inter-product agreement is strongest. The paper then extrapolates this urban-derived bound to non-urban vegetated environments where the noise environment is known to be considerably worse. This extrapolation direction is counterintuitive: if anything, the urban 95th-percentile figure represents a lower bound on uncertainty in vegetated settings, not an appropriate central estimate.

The threshold derivation has a second, more fundamental problem: it conflates disagreement between two methodologically dissimilar products with measurement uncertainty. O24 and W24 differ across four independent dimensions: sensor configurations (ALOS-1 + Sentinel-1 + GNSS vs. Sentinel-1 only), 3D displacement inversion vs. 1D line-of-sight-to-vertical conversion, temporal coverage (2007–2020 vs. 2017–2020), reference frame strategies, and native spatial resolution. Their disagreement at any given pixel, therefore, reflects a mixture of true observational differences (including legitimate temporal and spatial variability in SEC), methodological choices, and measurement noise. Treating the full inter-product spread as an uncertainty envelope assigns the consequences of methodological divergence equally to both products, regardless of which is closer to the ground truth.

The most acute problem, however, is internal inconsistency. The GNSS-derived background VLM rate that anchors the paper's GIA analysis, approximately −1.2 mm yr^{-1}, lies well within the range the paper simultaneously declares unresolvable. If InSAR-derived rates below 5 mm yr^{-1} are to be treated with "utmost caution," the same epistemological standard should apply consistently. The paper does not explain why GNSS is exempt from this caution at the −1.2 mm yr^{-1} level while InSAR is not. The answer that GNSS achieves sub-millimeter-per-year precision under favorable conditions, whereas InSAR does not, is correct, but it is precisely the kind of instrument-specific uncertainty characterization that the paper fails to provide. A rigorous uncertainty analysis would separately quantify GNSS and InSAR precision at these signal amplitudes, rather than applying a single inter-product-disagreement-derived threshold across both measurement systems.

5. Characterizations of O24 Performance Inconsistent with the Evidence

Beyond the methodological concerns above, several statements in Li et al. (2026) characterize O24's performance more negatively than the paper's own results support. We summarize these in Table 1 below, along with the relevant O24 evidence and our assessment of likely impact. We note that classifying a claim as a "potential misrepresentation" reflects a technical assessment of consistency with the published record and is not an allegation of intent.

Table 1. *Summary of statements in Li et al. (2026) that are inconsistent with O24 evidence or with the paper's own analysis, with classification and likely impact on their conclusions.*

Li et al. (2026) Statement	O24 Evidence	Classification	Technical Concern	Impact on Conclusions
Negligible spatial correlation between datasets ($R^2 = 0.05$).	O24 was developed and validated at 50 m resolution using GNSS-constrained inversion.	**Major methodological flaw**	O24 was aggregated 50 m → 1 km prior to comparison. This ~400× reduction in pixel density is not a neutral transformation.	May substantially exaggerate inter-product disagreement and understate O24 performance.

InSAR is 'only reasonably robust in densely urbanized settings.'	O24 was validated broadly along the U.S. coast and applied beyond dense urban areas with low reported errors.	**Potential misrepresentation**	The analysis does not isolate InSAR limitations from confounds: differing sensors, temporal windows, referencing strategies, and spatial support.	May lead readers to incorrectly conclude that coastal InSAR is broadly unreliable outside cities.
"Neither InSAR dataset fully captures the background VLM rate."	O24 demonstrates site-level GNSS correlation with RMSE < 1 mm yr^{-1} — acknowledged in the paper itself.	**Potential misrepresentation**	The authors' own results show measurable agreement; 'does not fully capture' mischaracterizes imperfect agreement as failure.	Overstates the negative interpretation of the comparison.
"InSAR data are presently unable to capture this rate."	O24 validation demonstrates sensitivity to mm/yr-scale deformation signals.	**Potential misrepresentation**	The evidence shows imperfect agreement, not inability. The claim is stronger than the data support.	May incorrectly imply a fundamental failure of InSAR methodology.
GNSS subset (110 → 36 → ~20 sites) used to establish background VLM benchmark.	O24 validation used a substantially broader GNSS framework across the full study domain.	**Major methodological flaw**	Progressive filtering excludes all Texas sites and concentrates the network on atypical stable upland locations.	Background VLM benchmark may not be representative of the broader coastal region.
5 mm yr^{-1} uncertainty threshold recommended for non-urban InSAR.	O24 uncertainty estimates were derived from independent GNSS validation, not inter-product disagreement.	**Major methodological flaw**	Threshold is derived from disagreement between two methodologically dissimilar products rather than from a rigorous uncertainty model.	May substantially overestimate uncertainty, leading to unwarranted discounting of valid observations.
Inter-product comparison interpreted as a reliability test.	O24 and W24 differ in sensors (ALOS-1 + Sentinel-1 vs. Sentinel-1 only), methodology, temporal coverage (13 vs. 4 years), referencing strategy, and resolution.	**Major methodological flaw**	Comparison conflates methodological and framework differences with accuracy. Disagreement may reflect legitimate observational differences, not error.	May attribute substantive measurement differences to error rather than to differing measurement frameworks.

Taken together, these inconsistencies suggest a systematic tendency in the paper to interpret ambiguous or mixed evidence in the direction most unfavorable to InSAR. Where O24's (and W24's) own results demonstrate measurable GNSS agreement and low site-level RMSE, the paper characterizes this as "failure to capture" rather than "imperfect agreement." Where inter-product differences could reflect methodological divergence rather than error, the paper attributes the full spread to uncertainty. These framing choices shape the paper's policy-facing conclusions in ways that go beyond what the data demonstrate.

6. Re-evaluating O24 and W24 Performance within Li et al. (2026)'s Study Area

To provide an independent assessment of O24 performance, we obtained daily vertical displacement time series in the IGS14 reference frame at 88 GNSS stations from the Nevada Geodetic Laboratory (NGL; https://geodesy.unr.edu/) within the Li et al. (2026) study domain. Following standard processing procedures, including removal of offsets, outliers, and common-mode errors, we estimated vertical velocities and associated uncertainties for two periods: 2007–2020 (corresponding to the O24 observational window) and 2017–2020 (corresponding to W24).

Figure 1 compares these period-specific GNSS velocity estimates. The mean difference between the two periods is −0.3 mm yr^{-1}, with a 95th-percentile absolute difference of 3.7 mm yr^{-1}. Applying Li et al.'s own logic to this result is instructive: if inter-measurement spread at the 95th percentile defines a caution threshold, as the authors argue in deriving their 5 mm yr^{-1} recommendation, then the equivalent threshold derived from this GNSS-to-GNSS comparison is 3.7 mm yr^{-1}, below which the authors' own GIA estimate of −1.2 mm yr^{-1} falls unambiguously. This reductio ad absurdum illustrates that the 5 mm yr^{-1} threshold, far from being a principled uncertainty bound, is an artifact of conflating inter-product methodological divergence with measurement error. No physically meaningful caution threshold can be derived by this approach without simultaneously invalidating the benchmark it is meant to protect.

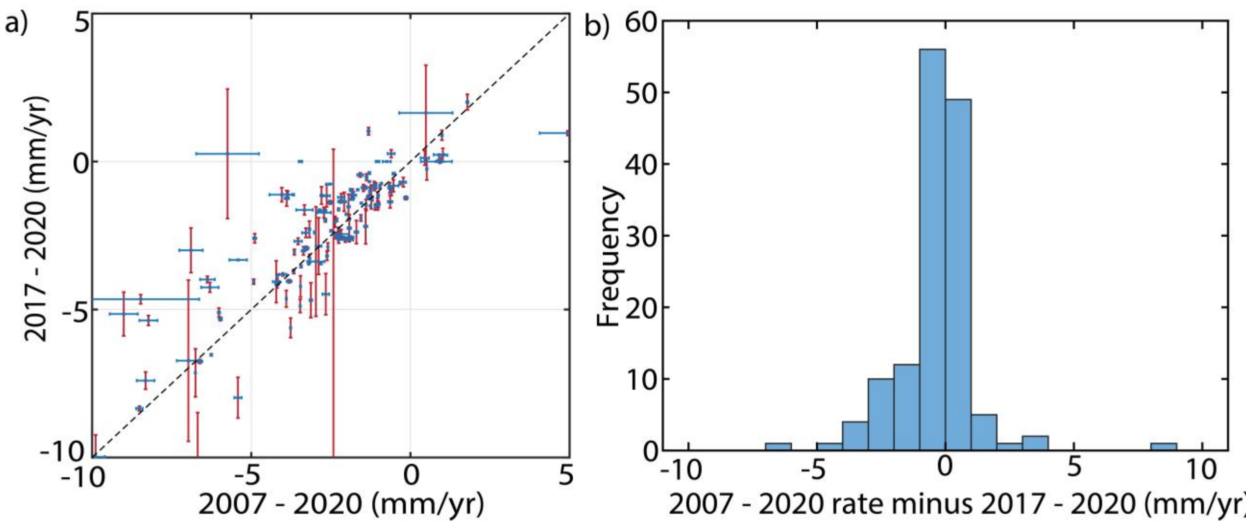

Figure 1. *Comparing GNSS rate for the 2007-2020 period vs the 2017-2020 period*

To assess O24 accuracy directly, we identified all O24 pixels within a 50 m radius of each GNSS station and computed the median InSAR-derived VLM value at each site. Figure 2 shows the resulting comparison between O24 and GNSS rates across the study domain. The standard deviation of the residuals is 1.6 mm yr^{-1} (mean residuals of -0.1 mm yr^{-1}), closely consistent with the 1.5 mm yr^{-1} validation uncertainty reported in Ohenhen et al. (2024) and well within the uncertainty bounds acknowledged by Li et al. (2026) for the original 50 m product. These results affirm that O24, evaluated at its native resolution and against a spatially representative GNSS network, performs robustly across the study area, in direct contrast to the characterization offered by Li et al. (2026).

We extended this validation to W24 over their 1km native resolution and 2017–2020 temporal window using the closest GNSS station to each pixel (Figure 3), obtaining a residual standard deviation of 2.3 mm yr^{-1} (mean residuals 1.1 mm yr^{-1}). When each InSAR product is instead validated against the GNSS temporal window it does not share, performance degrades. Validating W24 against GNSS rates over the 2007–2020 period raises the residual standard deviation from 2.3 to 2.8 mm yr^{-1} (a ~22% increase) and shifts the mean difference to 1.6 mm yr^{-1}; conversely, validating O24 against the 2017–2020 period raises its residual standard deviation from 1.6 to 1.9 mm yr^{-1} (a ~19% increase) and shifts the mean difference to

−0.6 mm yr⁻¹. Each product, therefore, agrees most closely with GNSS over its own observational window and degrades against the mismatched temporal window.

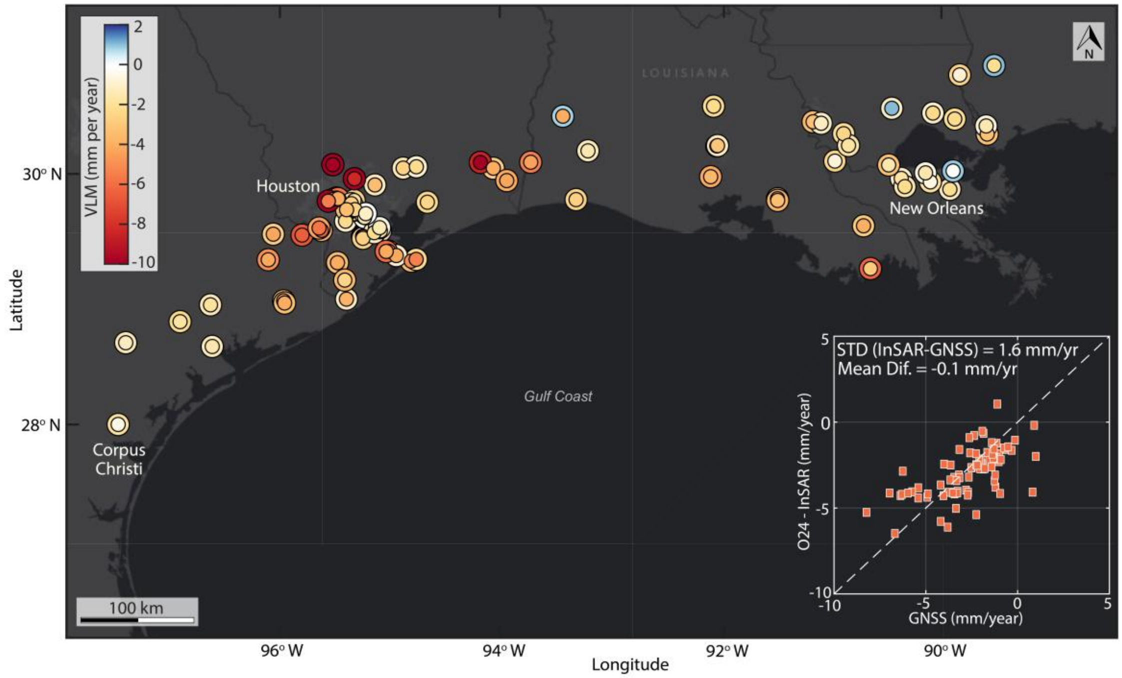

Figure 2. *Comparison of Ohenhen et al. 2024 (O24) rate (filled circles) against GNSS vertical rate (open circles) for the 2007-2020 period. The inset shows a bivariate plot comparing GNSS vertical rates with O24 InSAR-derived rates. The standard deviation of the difference between the two datasets is 1.6 mm/year.*

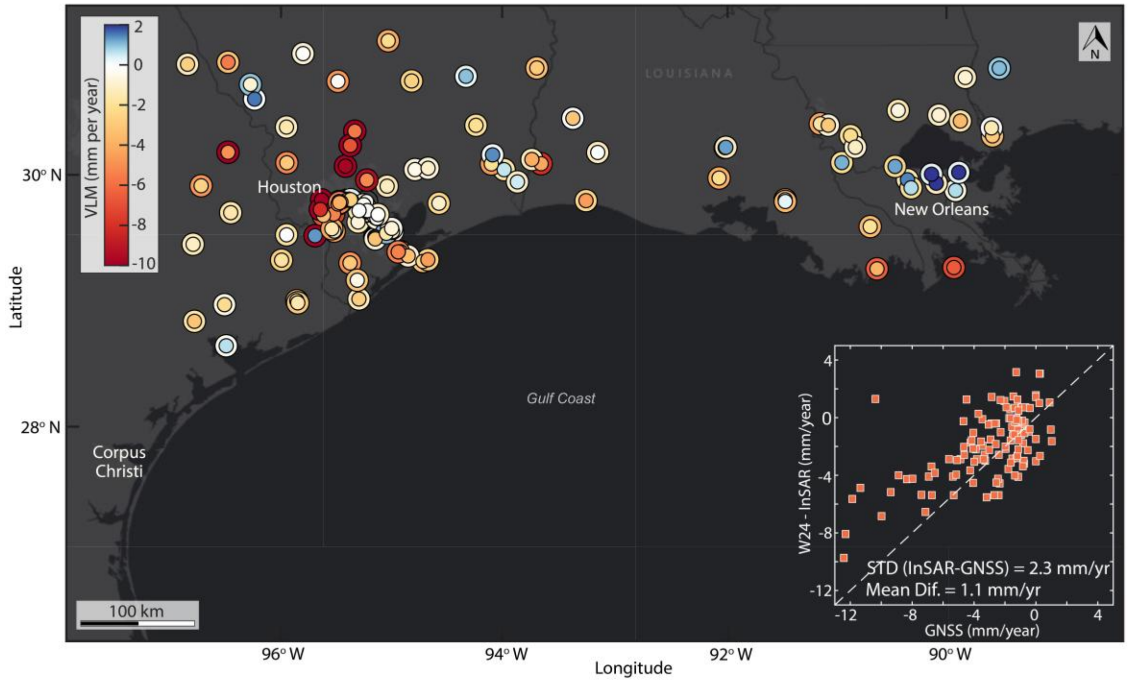

Figure 3. *Comparison of Wang et al. 2024 (W24) rate (filled circles) against GNSS vertical rate (open circles) for the 2017-2020 period. The inset shows a bivariate plot comparing GNSS vertical rates with W24 InSAR-derived rates. The standard deviation of the difference between the two datasets is 2.3 mm/year.*

7. Concluding Remarks

Li et al. (2026) address a legitimate and important question about InSAR reproducibility in vegetated coastal settings. Their call for harmonized processing frameworks, transparent methodological documentation, and systematic cross-validation reflects sound scientific priorities, and we support these recommendations unreservedly.

However, the specific conclusions of the paper that InSAR is presently unreliable for SEC rates below 5 mm yr^{-1} and that coastal subsidence maps should not yet be used for policy guidance, rest on an analysis that compares O24 at approximately 400 times coarser than its native resolution, validates against a small and geographically narrow network of stable-upland GNSS sites, applies an urban-derived uncertainty bound to non-urban environments, and characterizes O24 performance more negatively than the paper's own results justify. These are not minor technical details: each represents a choice that shifts the analysis in the same direction, toward a more pessimistic assessment of InSAR capability. Critically, our independent validation of O24 within the Li et al. (2026) study area, conducted at native 50 m resolution against 88 GNSS stations drawn from the full NGL network, yields a residual standard deviation of 1.6 mm yr^{-1}, consistent with the uncertainty reported by Ohenhen et al. (2024) and directly contradicting the paper's characterization of O24 as unable to resolve background VLM rates. When evaluated on its own terms, O24 (and W24) performs robustly; the apparent failure documented by Li et al. (2026) is, in substantial part, a consequence of the analytical choices applied to it.

We further demonstrate that the 5 mm yr^{-1} caution threshold is not a principled uncertainty bound but an artifact of conflating inter-product methodological divergence with measurement error. Applying the same logic to a GNSS-versus-GNSS period comparison within the study area yields a threshold of 3.7 mm yr^{-1}, below which the paper's own GIA estimate of −1.2 mm yr^{-1} falls, rendering it, by the authors' reasoning, equally unreliable. This internal contradiction exposes the threshold for what it is: a framework that cannot consistently adjudicate between signal and noise, and that should not be the basis for policy-facing conclusions about the state of coastal InSAR.

The appropriate response to genuine concerns about InSAR reproducibility is not a blanket caution threshold derived from a methodologically mismatched comparison, but rather a commitment to robust analyses that can distinguish between competing sources of inter-product disagreement. A comparison conducted at native resolution, with a geographically representative validation network, and with uncertainty estimates grounded in instrument-specific noise characterization rather than inter-product spread, would either substantiate Li et al.'s conclusions or reveal that InSAR performance in vegetated coastal settings is considerably more nuanced than the current analysis suggests. Either outcome would constitute a more durable and credible contribution to the field. The InSAR community would be better served by investing in that rigorous intercomparison infrastructure than by accepting artificially conservative uncertainty bounds that risk discrediting a methodology, and the policy-relevant datasets it has produced, on the basis of a flawed analytical design.

8. Acknowledgments

This work was supported by the U.S. Department of Defense. We thank Roland Burgmann and Robert Nicholls for their valuable comments and insightful discussions.

References

Li, G., Törnqvist, T. E., and Chen, J.: Evaluating InSAR-derived rates of surface-elevation change along the central U.S. Gulf Coast, Earth Obs., 1, 1–13, https://doi.org/10.5194/eo-1-1-2026, 2026.

Ohenhen, L. O., Shirzaei, M., Ojha, C., Sherpa, S. F., and Nicholls, R. J.: Disappearing cities on US coasts, Nature, 627, 108–115, https://doi.org/10.1038/s41586-024-07038-3, 2024.

Wang, K., Chen, J., Valseth, E., Wells, G., Bettadpur, S., Jones, C. E., and Dawson, C.: Subtle land subsidence elevates future storm surge risks along the Gulf Coast of the United States, J. Geophys. Res.-Earth Surf., 129, e2024JF007858, https://doi.org/10.1029/2024JF007858, 2024.